DOCUMENT RESUME

LI 000 991

ED 022 516

By-Buckland, M.K.; Woodburn, I.

AN ANALYTICAL APPROACH TO DUPLICATION AND AVAILABILITY. UNIVERSITY OF LANCASTER LIBRARY OCCASIONAL PAPERS, NO. 2.

Lancaster Univ. (England).
Spons Agency-Office for Scientific and Technical Information, London (England).
Pub Date Jun 68
Note-26p.
EDRS Price MF-$0.25 HC-$1.12
Descriptors-*COLLEGE LIBRARIES, *DECISION MAKING, INFORMATION UTILIZATION, *LIBRARY MATERIALS, MATHEMATICAL APPLICATIONS, MATHEMATICAL MODELS, OPERATIONS RESEARCH, PREDICTION, *UNIVERSITY LIBRARIES, USE STUDIES
Identifiers-*Poisson distribution

As part of a project designed to construct a mathematical model of the operation of an academic library, a research team determined the usefulness of a quantitative approach to making decisions about the number of copies of each item that a university library should provide. Using the reserve collection of the University of Lancaster Library as a basis, a mathematical relationship was established between the number of requests, the length of the loan period, the number of copies, and a standard of service (defined as "immediate availability"). It was found that the Poisson distribution could justifiably be used to relate the average request rate to the number of copies and the degree of availability likely to be achieved. A table was produced which can be used in a predictive manner insofar as the request rate itself can be predicted. A method of estimating the average request rate from incomplete data was also devised. It was concluded that a mathematical relationship between the factors concerned can be established. The chief difficulty lies in predicting the level of demand for library services and, more generally, in understanding the factors which determine it. (JB)

University of Lancaster Library Occasional Papers, No. 2

AN ANALYTICAL APPROACH TO DUPLICATION AND AVAILABILITY

by

M. K. BUCKLAND and I. WOODBURN

LANCASTER
1968

University of Lancaster Library Occasional Papers, No. 2

AN ANALYTICAL APPROACH TO DUPLICATION AND AVAILABILITY

by

M. K. BUCKLAND and I. WOODBURN

LANCASTER
1968

PREFACE

This research report is the second of a series of papers which the Library intends to publish at irregular intervals. It presents, in preliminary form, one aspect of a research project which has been supported since January 1967 by the Office for Scientific and Technical Information; this project is designed to construct a mathematical model of the operation of an academic library, in order that managerial decisions may be made on a rational rather than on an intuitional basis.

The research team comprises Mr. Ian Woodburn, M.A., and Mr. Michael K. Buckland, B.A., A.L.A.. Further reports will be issued in the coming months.

A. Graham Mackenzie

Librarian and Principal Investigator

June 1968

CONTENTS

SUMMARY: How many copies of each item ought a university library to provide? The authors attempt to answer this question using as their basis a reserve collection in the University of Lancaster Library. A mathematical relationship is established between the number of requests, the length of the loan period, the number of copies and a standard of service (defined as 'immediate availability'). The use of the Poisson distribution is found to be justified. A method of estimating total demand is suggested, using only the records of satisfied requests. The chief difficulty is seen to be not the calculation of the minimum number of copies to meet a given demand, but the prediction of what the level of demand is likely to be.

1. INTRODUCTION

1.1. THE PROBLEM OF DUPLICATION

A major difficulty in libraries is deciding how many copies of a given book should be made available for students. Even when the librarian knows which titles are likely to be in demand, there remains the problem of estimating how many copies of each will be needed. If too few are provided, then library users will be unable to find the documents which they need, and the library is failing in one of its fundamental tasks; if too many copies are provided, then money has been wasted and the library's budget is being inefficiently used, in that the cost of buying, processing and storing the redundant copies could have been more profitably used for other purposes. This problem is especially acute where a number of students are taking the same or similar courses.

It would, therefore, seem convenient if an examination of the various factors concerned could provide guidance in this matter. In particular we have been interested in establishing, for any given item, the relationship between:

i). The number of requests,

ii). The loan period,

iii). The number of copies, and

iv). A standard of service.

It is, in fact, convenient to combine the first two of these factors: if the number of requests occurring during any given span of time is divided by the number of loan periods in the same span of time, then we have an "average request rate per loan period". The effect on the average request rate of an increase or decrease in the number of requests will be clear. So too is the effect of a change in the loan period: to lengthen the loan period is to reduce the number of loan periods in any given span of time, and so the average request rate per loan period is increased; conversely the shortening of the loan period decreases the average request rate per loan period. This report uses the concept of an average request rate throughout, but it should not be forgotten that the effect of a change in the length of the loan period can readily be assessed by calculating a new request rate.

1.2. THE CONTEXT OF THIS REPORT

Although the work to be described may well be applicable in many

other libraries, it must be stressed that we have not undertaken it merely as a theoretical exercise. This report is, in effect, an interim account of an attempt to see whether, and how far, an analytical approach could be used to guide day-to-day decisions about duplication in an actual reserve collection within the University of Lancaster Library.

This reserve collection is known as the Short Loan Collection: all material known, or expected, to be in heavy demand is removed from the open shelves to this collection which is located, on closed access, behind the service desk. This means that a reader wanting an item from the collection has to ask for it and, if it is available, a member of the library staff will issue it.

With a view to increasing the number of readers who can use each copy, and thus reducing duplication, the loan period is short - normally about four hours. There are twenty-two such loan periods in a week: each weekday has four (morning, afternoon, evening and overnight), and Saturday has two (morning, and from noon until Monday morning). At whatever time an item is borrowed, it must be returned or re-borrowed by the end of the same loan period; this rule is strictly enforced by a scale of heavy fines.

In general the teaching staff are relied upon to give prior notice of books which are likely to be heavily used, and of the appropriate number of copies of each, since it is they who direct the student's reading.

2. THE POISSON DISTRIBUTION

2.1. BACKGROUND

Although it may well be possible, by keeping careful records, to establish the average number of requests per loan period for a given title, it does not follow that precisely this number of requests will be made in each individual loan period. In fact, the actual number of requests appears to fluctuate somewhat haphazardly from one individual loan period to another.

Statisticians have often found, when examining this sort of situation, that in the long term there is a predictable regularity in the extent of such fluctuations, which can be described by a mathematical formula known as the Poisson distribution.

$$P(s) = \frac{e^{-r} \quad r^{s}}{s!}$$

This states the probability P (s) of s requests being made in a single loan period, where r is the average request rate and e is a constant (approximately 2.71828).

For example, if the average request rate were one request per loan period, then we should not expect that there would be precisely one request in each. Instead we should expect that although there often was one request in a loan period, there would sometimes be none, two, or three, and on rare occasions four or more. The Poisson distribution indicates that if the average request rate is one per loan period, then the relative frequency of 0, 1, 2, 3, etc., requests occurring corresponds to the height of the columns in figure 1.

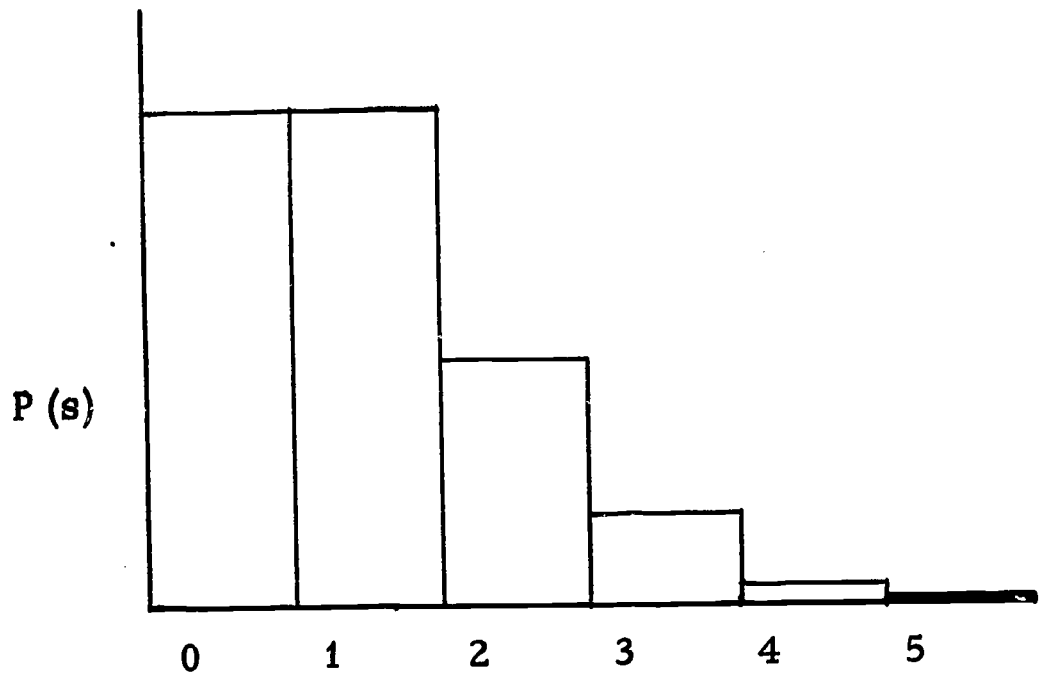

FIGURE 1.

This formula has been used in many different fields; the outstanding example of its use in librarianship is the planning of the National Lending Library[1]. It has also figured in other studies (e.g. refs. 2-5).

2.2. ESTABLISHING A RELATIONSHIP BETWEEN DUPLICATION, AVAILABILITY AND THE PATTERN OF REQUESTS

The objective of this analysis is to derive a relationship between:

(1) Request pattern,

(2) Number of copies,

(3) Immediate availability.

In investigating the request pattern, we assume initially that there is an ascertainable average request rate for a given title, and that the Poisson distribution can adequately describe the variations in the observed numbers of requests in individual loan periods.

The concept of immediate availability is based on the simple fact that when an item is requested it is either available there and then, or else it is not. The immediate availability rate is the

proportion of requests which are immediately satisfied, and can conveniently be stated as a percentage:

$$\% \text{ availability} = \frac{100 \times \text{no. of requests immediately satisfied}}{\text{total no. of requests}}$$

This concept is admittedly a somewhat crude measure of the standard of service; for example, it does not take fully into account the fact that requests which are not immediately satisfied can almost invariably be satisfied on some later occasion if the user has sufficient time and persistence. However, such delays are prima facie undesirable, and in the circumstances immediate availability seemed to us to be a reasonable measure of the standard of service. Accordingly throughout this report the terms "available" and "satisfied" are used to mean immediately available and immediately satisfied.

In order to establish the availability rate we need to know both the number of satisfied requests and the total number of requests. If the number of requests for a title in one loan period is equal to, or less than, the number of copies available, then all requests will be satisfied. If the number of requests exceeds the number of copies, and if, as we assume, a copy cannot be borrowed more than once within a single loan period, then the number of satisfied requests cannot exceed the number of copies.

We define n as the number of copies of a given title and s as the number of requests for it in any single loan period so that when $0 \leqslant s \leqslant n$ then s requests are satisfied and when $s > n$ then n requests are satisfied.

Since P (s) expresses the relative frequency of s requests being made in a loan period

$$\% \text{ availability} = \frac{100 \left[\sum_{s=0}^{s=n} s\,P(s) + \sum_{s=n+1}^{s=\infty} n\,P(s) \right]}{\sum_{s=0}^{s=\infty} s\,P(s)}$$

We now have a relationship involving:

i). The number of requests,

ii). The loan period,

iii). The number of copies, and

iv). % availability.

From this we can calculate the dependence of the availability rate on such factors as

a). the length of the loan period, and

b). the number of copies in stock.

In the case of the Short Loan Collection a reduction in the length of loan period has been considered undesirable, and our attention has, therefore, been focussed on predicting the correct number of copies.

Example. Average request rate of 1 request per loan period.

No. of copies	Availability
1	63.2%
2	89.6%
3	97.7%
4	99.6%
5	99.9%

TABLE 1.

This information can be summarised more conveniently in the following table.

Average Request rate	Minimum no. of copies required for 80%	90%	95% availability
0.1	1	1	1
0.2	1	1	2
0.3	1	2	2
0.4	1	2	2
0.5	2	2	2
1.0	2	3	3
1.5	2	3	4
2.0	3	4	4
2.5	3	4	5
3.0	4	5	5
3.5	4	5	6
4.0	4	6	6

TABLE 2.

We would stress that this approach is best suited to comparatively simple situations such as the Short Loan Collection being investigated. In a typical open access library, items are used in at least two ways: they are consulted inside the library for short periods and they are borrowed for comparatively long and widely varying periods. Consequently, the librarian is likely to be especially interested in possible variations in loan policies, and in the interplay of two or more kinds of use. (For example, how far can he afford to allow the borrowing of material which is traditionally confined to the library?) Furthermore, in such a situation, the librarian is - understandably - concerned with the effects of various and sometimes expensive arrangements for reservations and recalls, and with the delays involved in meeting requests which have not been satisfied immediately. For this more complex situation we are developing a different approach which will be described in a future report.

2.3. IS THE USE OF THE POISSON DISTRIBUTION JUSTIFIABLE?

The results in Section 2.2 were based on the hypothesis that the pattern of requests predicted by the Poisson distribution would be similar to the actual pattern of requests in the collection concerned.

The ideal situation for testing this hypothesis would be where it was feasible to record all requests (not only satisfied requests) without the user's knowledge, and where the user could not tell in advance whether or not an item was available. The Short Loan Collection within the University of Lancaster Library corresponds closely to this ideal situation; consequently for a period of two weeks during May 1967 all requests were examined. Satisfied requests were already being recorded on the date labels in each item: the date stamped recorded the day of issue, and its position on the date-label the loan period within that day. Arrangements were made for the library staff to note briefly on a piece of paper the details of each request which was not immediately satisfied. Users are unlikely to have known in advance whether or not an item would be available, since it is not normally possible for them to see if there is still a copy of a desired item on the shelves; however, we cannot rule out the possibility that one user may occasionally have learned from another that all copies were already on loan. To the best of our knowledge none of the users was aware that data concerning their requests was being collected.

During these two weeks 2,010 requests were recorded, and the immediate availability rate for the collection as a whole was established as 90%. A small random sample of 25 titles was taken, and 13 different patterns of use emerged: in each case the total number of requests occurring during the fortnight was divided by the number of loan periods during the survey (44) in order to produce an average request rate per loan period for each title. The Poisson distribution was then used to predict a pattern of requests for each

average request rate. In all cases, except one, the differences between the observed and the predicted patterns are not statistically significant, as the chi-squared values in Table 3 indicate. In view of this similarity, we feel that the use of the Poisson distribution for this purpose is justified.

TABLE 3 - COMPARISON OF OBSERVED AND PREDICTED REQUEST PATTERNS

R = No. of requests. O = Observed frequency. P = Frequency predicted by the Poisson distribution.

R	O	P	O	P	O	P	O	P	O	P
0	15	14.12	23	24.37	26	28.57	29	29.90	35	33.50
1	13	16.05	16	14.40	17	12.34	13	11.55	7	9.14
2	11	9.12	5	4.25	1	2.66	2	2.23	1	1.25
3	5	3.45	0	0.84	0	0.38	0	0.29	1	0.11
4	0	0.98	0	0.12	0	0.04	0	0.03	0	0.008
5	0	0.22	0	0.01	-	-	-	-	-	-
6	0	0.04	-	-	-	-	-	-	-	-
7	0	0.007	-	-	-	-	-	-	-	-
		Chi-sq: 2.967		Chi-sq: 1.362		Chi-sq: 3.457		Chi-sq: 0.548		Chi-sq: 7.566
		DF: 7		DF: 5		DF: 4		DF: 4		DF: 4
0	31	32.74	33	32.74	34	35.05	40	40.18	37	35.86
1	13	9.67	9	9.67	10	7.97	4	3.65	5	7.34
2	0	1.43	2	1.43	0	0.90	0	0.17	2	0.75
3	0	0.14	0	0.14	0	0.07	0	0.005	0	0.05
4	0	0.01	0	0.01	-	-	-	-	-	-
		Chi-sq: 2.816		Chi-sq: 0.428		Chi-sq: 1.524		Chi.sq: 0.205		Chi-sq: 2.913
		DF: 4		DF: 4		DF: 3		DF: 3		DF: 3
0	43	43.01	42	42.04	42	41.10				
1	1	0.98	2	1.91	1	2.80				
2	0	0.01	0	0.04	1	0.10				
		Chi-sq: 0.012		Chi-sq: 0.048		Chi-sq: 9.742				
		DF: 2		DF: 2		DF: 2				

3. APPLICATION

3.1. PLANNING FOR FUTURE DEMAND

In Section 2.2. we demonstrated how a simple table could be produced to show the minimum number of copies required for different levels of availability and for various average request rates. If it is known in advance what the average request rate will be, then it is a very simple matter to discover how many copies will be needed to maintain any desired level of availability. Unfortunately, it may not be easy to predict the average request rate for any given title.

The information which could normally be made available to the librarian is the number of students who have been advised to read an item and the span of time during which they are all expected to read it (e.g. the date by which an essay should be completed.) If b students request an item once only during a period of time containing p loan periods, then the request rate can be written

$$r = \frac{b}{p}$$

It would be rash, however, to assume that each item will in fact be requested once and once only by each student. Some may not request it at all because they have access to another copy elsewhere, or because they simply do not bother. Others may make more than one request because one borrowing was insufficient or because the book was not available when first requested. These conflicting tendencies might cancel each other out but it may be better to assume that

$$r = x \cdot \frac{b}{p}$$

where x is a quantity, or parameter, which corresponds to this variation in user behaviour. Further investigation might reveal that the value of x tended to remain constant for a particular group

of students, or for a particular course, or even for a particular title.

In any case the people most closely concerned - the Service Desk staff and the teaching staff - should be able to estimate from their experience the number of requests for each item.

The main difficulty in the use of this approach for planning for future demand lies in predicting the average request rate. It is necessary, therefore, to explore the consequences of inaccurate prediction.

For example, if the average request rate is expected to be between 0.5 and 1.5 requests per loan period, then we can examine the likely consequences of different decisions about the number of copies to be provided. (See Table 4).

Average Request rate	Availability resulting from the provision of 1	2	3	4 copies
0.5	78.7%	96.7%	99.6%	99.96%
1.5	51.8%	81.3%	94.0%	98.4%

TABLE 4.

Similarly if the request rate could only be established as lying between 2 and 4:

Average Request rate	Availability resulting from the provision of 3	4	5	6 copies
2	89.1%	96.2%	98.9%	99.7%
4	66.3%	80.5%	89.7%	95.1%

TABLE 5.

3.2. ADJUSTING EXISTING PROVISION

The librarian may well wonder whether there are sufficient copies of a given item and, if not, how many additional copies there ought to be.

The number of satisfied requests for a given title is normally easy to discover from date labels or issue records, but he also needs to know the number of unsatisfied requests before he can establish the availability rate. If he considers that this is too low, then he has to decide how many additional copies are required. This can easily be done by reference to a table such as Table 2 in Section 2.2, provided that he knows the average request rate (defined as the total number of requests during a given span of time divided by the number of loan periods in the same span of time). Again he needs to measure, or at least estimate, the number of unsatisfied requests.

At the collection actually studied, it happened to be easy to record unsatisfied requests once it had been decided to do so. Nevertheless, the situation might well arise, at Lancaster or elsewhere, in which data on unsatisfied requests was needed retrospectively for a period during which it had not been collected; in different circumstances it may not be practicable to record the number of unsatisfied requests at all. In such cases it is necessary to rely on an estimate.

3.3. ESTIMATING THE REQUEST RATE FROM THE NUMBER OF SATISFIED REQUESTS ALONE

Assuming that data on satisfied requests is available, the problem of estimating the total number of requests in any period of time (and hence the request rate) is in the following form: let us imagine the case where there are three copies of a book and that we have tabulated the data on satisfied requests. If no requests were made during a loan period then there can have been no satisfied

requests, and we know from our data how often this has happened. Similarly when one or two requests were made, and there are three copies, then all requests will have been satisfied and, again, we know the frequency of this occurrence. So far we may be confident that the total number of requests and the number of satisfied requests were the same; but the trouble begins with the remaining loan periods in which three satisfied requests have been recorded. We only know that at least three requests were made; we do not know on how many of these occasions there were four or more. Yet this is the crucial information: three requests could be satisfied, but if there were more, then at least one request must have gone unsatisfied. What we really need to know is the number of occasions upon which four, five, six, etc., requests were made.

However, if we are justified in assuming that, for any given item, the fluctuations in the number of requests in individual loan periods correspond to the type of fluctuations predicted by the Poisson distribution, it is possible to estimate, from the incomplete data which we have, what the average request rate is most likely to have been. Furthermore it is also possible to gain an impression of the amount of confidence we are entitled to place in the answer.

Since the available data includes only satisfied requests, it can be tabulated as follows where n is the number of copies.

No. of requests per loan period	Observed frequency
0	f_0
1	f_1
2	f_2
⋮	⋮
n-1	f_{n-1}
$\geqslant n$	f_n

TABLE 6.

i.e. we can only observe the number of occasions f_n when n or more requests were made.

If we assume that in any given loan period the actual number of requests will follow a Poisson distribution, then the likelihood of observing the tabulated data is given by

$$L = \prod_{s=0}^{s=n-1} \left(\frac{e^{-r} r^s}{s!}\right)^{f_s} \cdot \left[1 - \sum_{s=0}^{s=n-1} \frac{e^{-r} r^s}{s!}\right]^{f_n}$$

$$\therefore \log L = \sum_{s=0}^{s=n-1} f_s \log \left(\frac{e^{-r} r^s}{s!}\right) + f_n \log \left(1 - \sum_{s=0}^{s=n-1} \frac{e^{-r} r^s}{s!}\right)$$

If we take a series of possible values of r, the request rate, then the corresponding values of L will indicate which value of r (the average request rate) is most likely, and will also give some indication of the extent to which this value is more likely than other possible values. The reliability of this method has not yet been examined.

Example.

Data

No. of requests per loan period	Observed frequency
0	29
1	13
⩾2	2

Result

r	Log L
0.35	- 34.39303
0.37	- 34.31419
→ 0.39	- 34.28505 ←
0.41	- 34.30062
0.43	- 34.35666

TABLE 7.

We conclude that 0.39 is the request rate most likely to result in the data observed.

It would also seem reasonable to attempt to apply this approach to the special case of gross underprovision (when all copies are borrowed every loan period): What is the lowest request rate which would produce this effect? Consequently what is the lowest number of additional copies required?

Example.

Data

No. of requests per loan period	Observed frequency
0	0
1	0
$\geqslant 2$	44

Results

r	Log L
2	- 22.91898
4	- 4.22604
6	- 0.77016
8	- 0.13304
10	- 0.02198

TABLE 8.

It would be reasonable to conclude that, in this hypothetical example, the request rate is likely to have been at least 6 requests per loan period and that the number of copies provided should be increased accordingly.

3.4. FURTHER APPLICATIONS

The entire analysis has been concerned with individual items, and the problem has been regarded as one of ensuring that there is

sufficient duplication for each item to reach a specified level of availability.

A librarian might wish to tackle rather different problems, and the approach outlined in this report could still be useful. For example, if he cannot afford to bring all titles in his stock up to a specified level of availability, then his problem becomes one of selecting the titles of which extra copies will bring the greatest benefit to readers. Since the individual items are normally considered separately, the consequences of buying additional copies of any two or more items can be compared, so that the final selection results in a greater increase in satisfied requests than any other possible selection.

Although the analysis has been concerned with individual titles, it is important to remember that no distinction is made between items having the same average request rate. Consequently, the items in a given collection can be grouped according to their request rates. By considering each group separately and then aggregating the results, it would be possible to answer such questions as: What would be the effect on the collection as a whole of a change in the length of loan period?

4. CONCLUSIONS

We originally began to examine the Short Loan Collection at the University of Lancaster Library with two aims: Firstly, how well was it working? Secondly, how far could a quantitative approach be helpful in taking routine decisions about the number of copies to be provided?

When we realised that the unsatisfied requests could conveniently be recorded, the availability rate was determined and discovered to be somewhat higher than we had subjectively expected. The next stage was to develop a mathematical relationship for individual items, between

i). The number of requests,

ii). The loan periods,

iii). The number of copies, and

iv). The degree of immediate availability, which had been chosen as the standard of service.

The number of requests and the number of loan periods were related to produce an average request rate. The assumption was made, and subsequently tested, that the Poisson distribution could justifiably be used to relate the average request rate to the number of copies and the degree of availability likely to be achieved. Consequently a table has been produced which can be used in a predictive manner insofar as the request rate itself can be predicted. It can also be used in a corrective manner, but, of course, any corrective action would have to be taken promptly else the circumstances might change.

In order to assist the librarian a method of estimating the average request rate from incomplete data was devised. No attempt has been made to indicate what level of availability the librarian

ought to endeavour to achieve; nor has a serious attempt yet been made to isolate and relate the factors determining the demand for library services; these factors might very well include the level of availability achieved. The analytical approach described in this report appears to be adequate for the circumstances investigated, but it might not be suitable in more complex situations.

The main conclusions are clear. A mathematical relationship between the factors concerned can indeed be established. The real difficulty lies elsewhere: in predicting the level of demand for library services and, more generally, in understanding the factors which determine it.

5. REFERENCES

1. URQUHART, D.J. and BUNN, R.M. A national loan service for scientific serials. Journal of Documentation 15 (1) March 1959, pp. 21-25.

2. DAMIRCHI, A.H. Problems in Operations Research: M.I.T. Science Library; submitted in partial fulfillment of the requirements for the degree of Bachelor of Science at the Massachusetts Institute of Technology. Cambridge, Mass., 1957.

3. LEIMKUHLER, F.F. Systems analysis in university libraries. College and Research Libraries 27 (1) Jan. 1966, pp. 13-18.

4. MORSE, P.M. Probabilistic models for library operations. In: ASSOCIATION OF RESEARCH LIBRARIES. Minutes of the 63rd meeting, January 26, 1964, Chicago. [Washington, ARL, 1964?]. Appendix A, pp. 9-19.

5. MORSE, P.M. On the prediction of library use. In: INTREX: report of a planning conference on information transfer experiments, ed. by C.F.J. Overhage and R.J.Harman. Cambridge, Mass., M.I.T. Press, 1965. Appendix N, pp. 225-234.

University of Lancaster Library Occasional Papers.

1. Some implications for library management of scattering and obsolescence, by M. K. Buckland and I. Woodburn. May 1968.

2. An analytical approach to duplication and availability, by M. K. Buckland and I. Woodburn. June 1968.